



# ***The US DOE Joint Genome Institute***

***Trevor Hawkins***  
***Director***

## Kelsey pics



## G5 and DOE goes to China



US DOE Joint Genome Institute



## JGI Experience

- **April 2000:** JGI drafted human chromosomes 5, 16 and 19 (11% of the genome)
- **October 2000:** JGI draft sequences 20 microbial genomes
- **June 2001:** JGI draft sequences White Rot Fungus
- **July 2001:** JGI sequences syntenic mouse regions of human chrom 19
- **October 2001** JGI completed Fugu and Ciona draft sequences

**Microbial**



**Fugu**



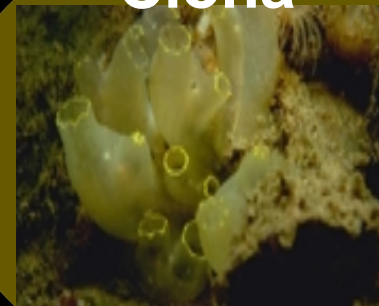
**HUMAN**



**White Rot Fungus**

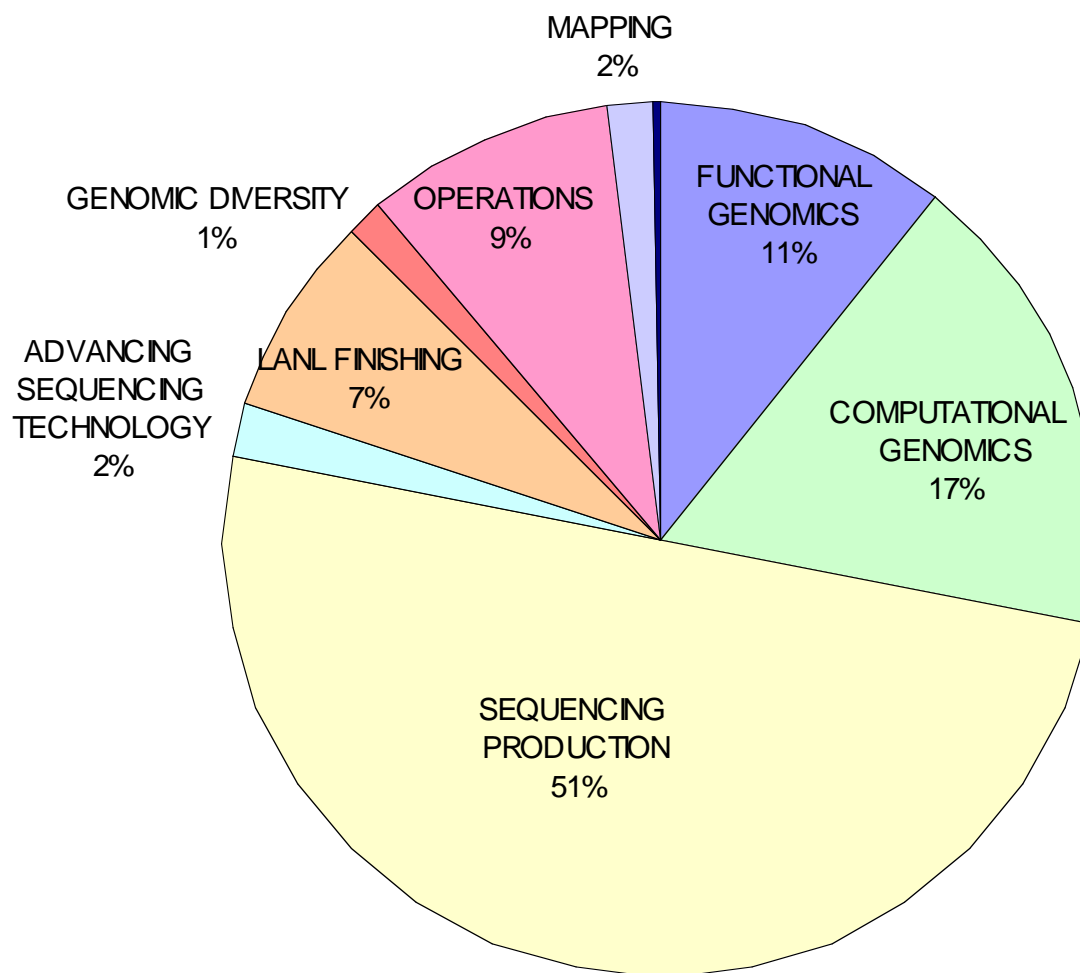


**Ciona**



## JGI FY02 Budget Distribution by Task

- TOTAL JGI Budget: \$77M
- DOE \$60M
- Grants ~\$5M
- In addition we fund Stanford human genome center (~\$5M FY02) to finish chromosomes 5 and 19.
- New initiatives ~\$7M



## JGI Production DNA Sequencing

- **JGI has one of the largest DNA sequencing operations in the world.**
- **Costs per phred 20 base are the key comparison tool**
- **JGI produce ~40 million phred 20 bases today.**
- **Planned increases to ~55 million phred 20 bases in February 2002**
- **Sequencing will utilize 384 capillary systems (MegaBACE 4000) by early next year**

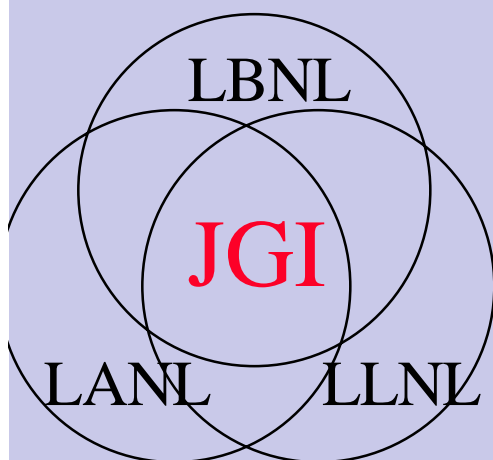
## An “Opportunity”



- Human genome sequencing
- Non pathogen microbial sequencing
- Computation & Analysis
- Functional Genomics



- Methods for forensic detection of pathogens
- BASIS early warning and detection system available for deployment to the field
- Partnership with CDC/ FBI, developed “gold standard” DNA assays for pathogen detection



## An “Opportunity”

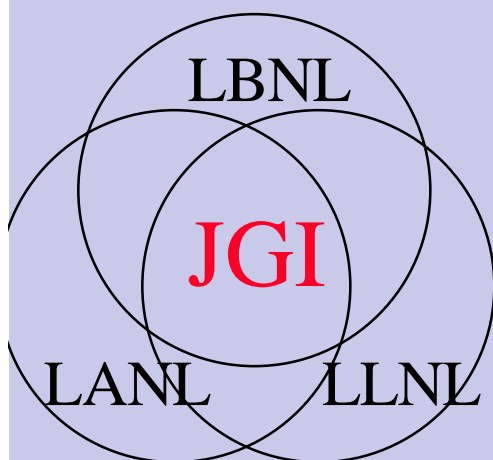
- Human genome sequencing
- Non pathogen microbial sequencing
- Computation & Analysis
- Functional Genomics



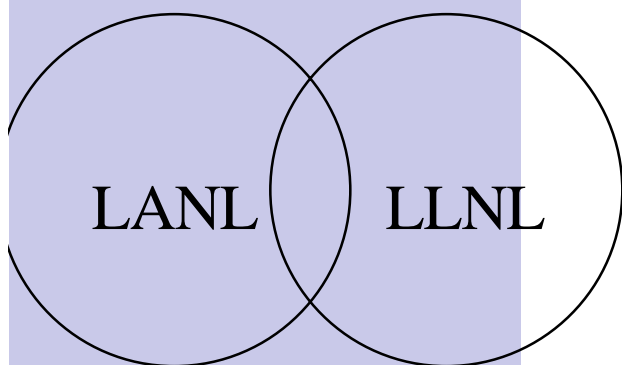
- Methods for forensic detection of pathogens
- BASIS early warning and detection system available for deployment to the field
- Partnership with CDC/ FBI, developed “gold standard” DNA assays for pathogen detection



## An “Opportunity”



- Human genome sequencing
- Non pathogen microbial sequencing
- Computation & Analysis
- Functional Genomics



- Methods for forensic detection of pathogens
- BASIS early warning and detection system available for deployment to the field
- Partnership with CDC/ FBI, developed “gold standard” DNA assays for pathogen detection

## Planned & new funding opportunity changes to the JGI production sequencing

	<u>Existing JGI February 2002</u>	<u>Microbial Initiative</u>
# Phred 20 bases/day	~55M	~20M
Cost / lane	\$2	\$2
Cost / Phred 20	\$0.0037	\$0.0037

### JGI Production Stats

Pass Rate: 90%

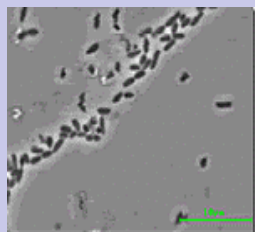
Read length: 600 bases

See [www.jgi.doe.gov](http://www.jgi.doe.gov)  
for latest results

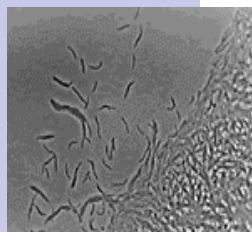
- Planned changes are use of new MB 4000 instruments & fulfill current JGI mission goals
- New funding additional lanes are through continued use of existing instruments
- Pass rate and read length should be factored in to comparing lane costs

# JGI Microbial Genomes

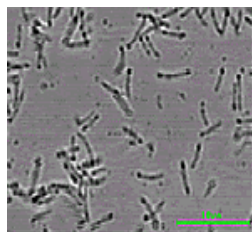
*A selection of genomes sequenced*



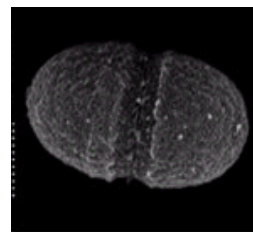
*Burkholderia  
cepacia*



*Cytophaga  
hutchinsonii*



*Desulfitobacterium  
hafniense*



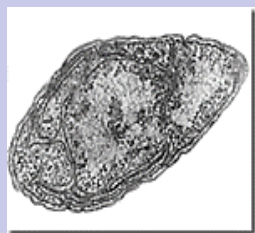
*Enterococcus  
faecium*



*Ferroplasma  
acidarmanus*



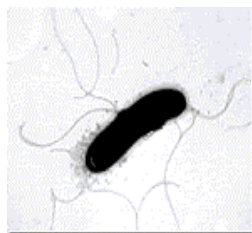
*Magnetospirillum  
magnetotacticum*



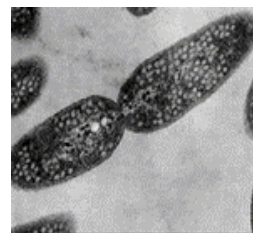
*Nitrosomonas  
europaea*



*Prochlorococcus  
marinus*



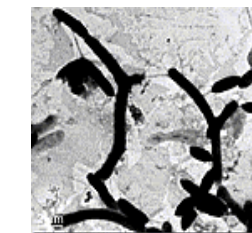
*Pseudomonas  
fluorescens*



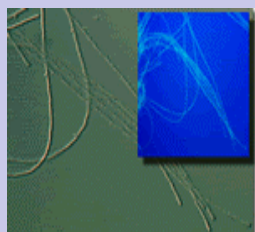
*Rhodobacter  
sphaeroides*



*Rhodopseudomonas  
palustris*



*Sphingomonas  
aromaticivorans*



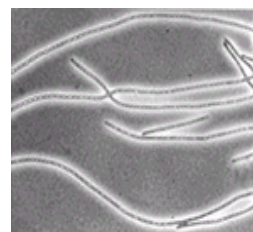
*Thermomonospora  
fusca*



*Trichodesmium  
erythraeum*



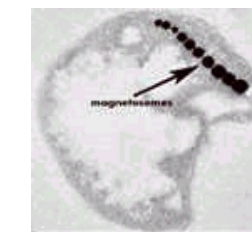
*Xylella  
fastidiosa*



*Nostoc  
punctiforme*



*Marine  
synechococcus*



*Magnetococcus  
MC-1*

## Microbial Genome Statistics

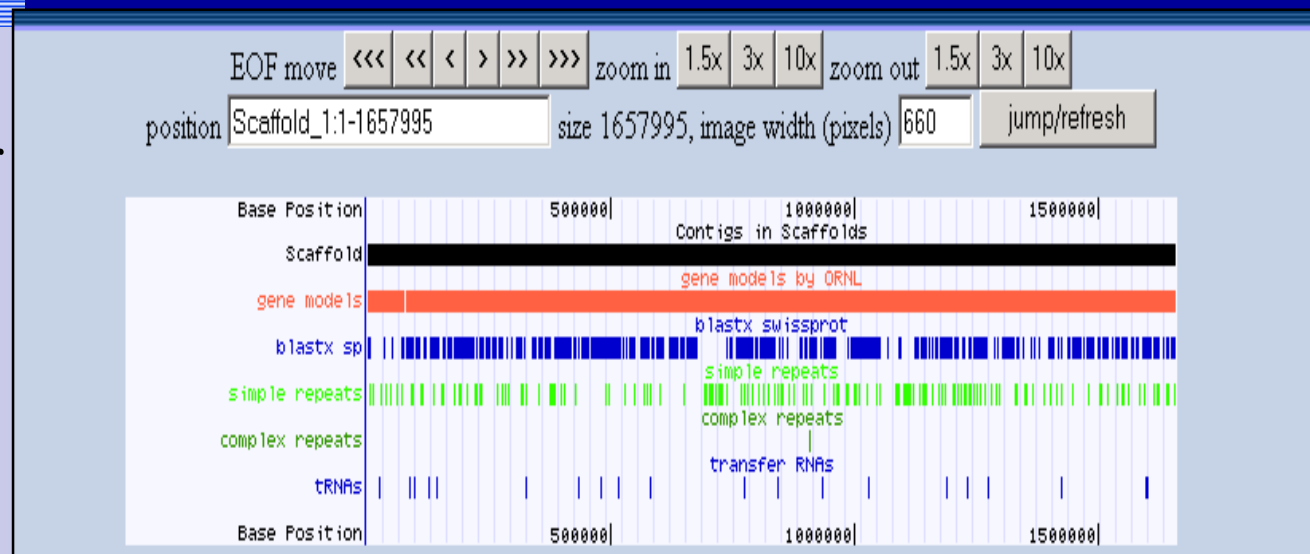
### •One Genome – One scaffold

- Fosmid libraries to be made for all
- 1 end/ 5Kb
- Directed lab work to link scaffolds

<u>Type</u>	<u>Size</u>	<u># Genomes</u>
In process	85 Mb	30
Drafted	72.4 Mb	16
Completed	14.8 Mb	5
<b>Total</b>	<b>172.2 Mb</b>	<b>51</b>

# Genome Browser *Prochlorococcus marinus*

Web based viewer  
used for all  
JGI genomes

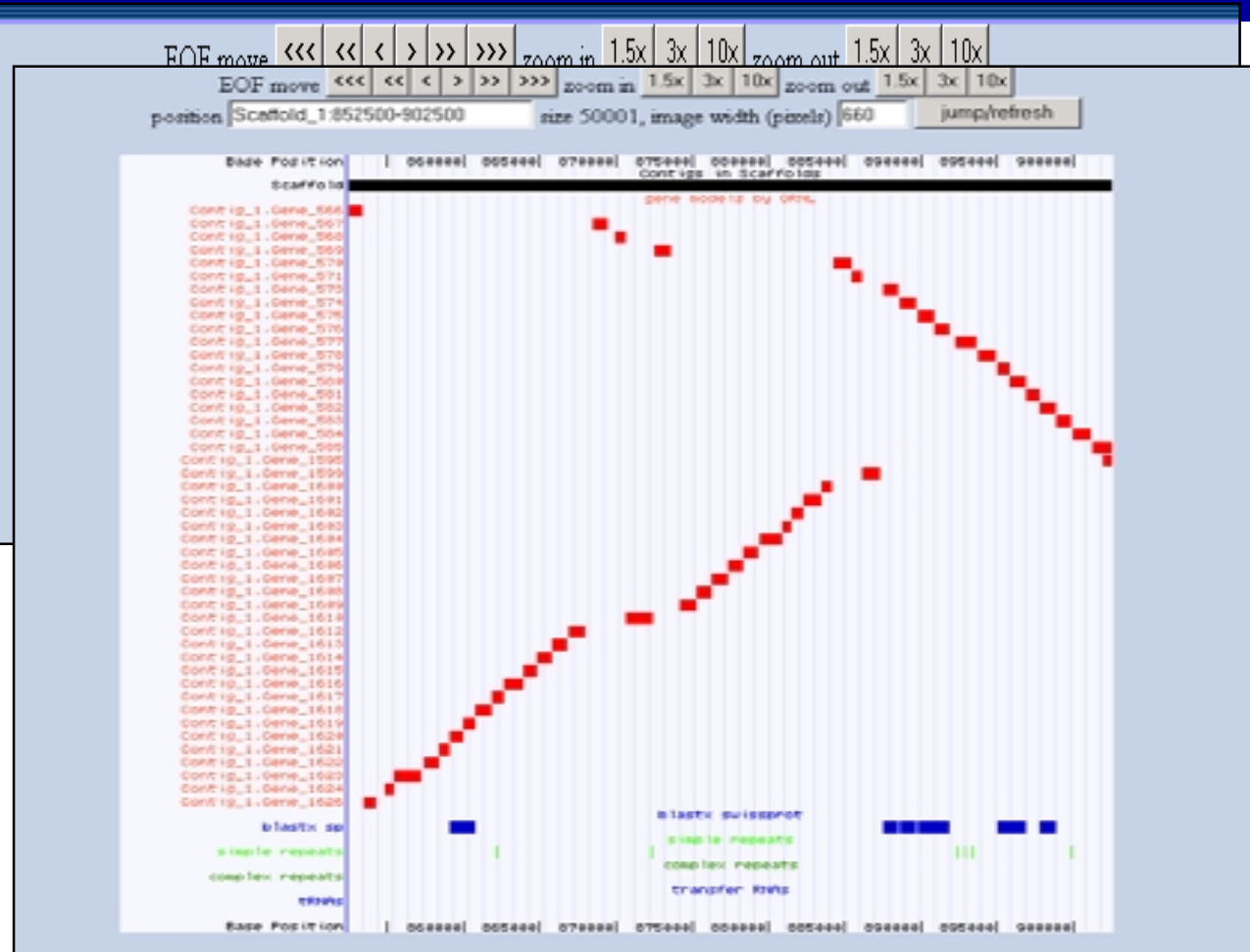




# Genome Browser *Prochlorococcus marinus*

Web based viewer  
used for all  
JGI genomes

Gene models as  
Provided by  
ORNL



# Gene models as Provided by ORNL

# Links to HMM & Protein homology hits



# Genome Browser *Prochlorococcus marinus*

Web based viewer  
used for all  
JGI genomes

Gene models as  
Provided by  
ORNL

Links to HMM &  
Protein homology  
hits

Links to PIR &  
Other online  
infomation

The screenshot displays the JGI Genome Browser interface for *Prochlorococcus marinus*. The top navigation bar includes links for 'JGI Home', 'Search P. marinus Med4', and 'Browse P. marinus Med4'. Below this, the 'Gene' section shows 'On Scaffold' and 'HMM H'. The 'Protein' section features a 'PIR Entry T44825' with the following details:

- ENTRY:** T44825
- TITLE:** Hypothetical protein woe (imported) - Acetabacteriota
- ALTERNATE\_NAMES:** protein tyrosine kinase
- ORGANISM:** #Formal\_name Acetabacteriota\_inoffici
- DATE:** 21-Jan-2000
- ACCESSION:** T44825
- REFERENCE:** 212814
- SEQUENCE:** 1 5 10 15 20 25 30

The sequence is displayed in a grid format, showing the amino acid sequence for the protein. The interface also includes a 'jump/refresh' button and a 'Scores' section.

## Data Release Policy

Version	Description	Delay After Draft Sequence
1	Un-edited assembly of small insert library	2 weeks
2	Hand-edited assembly of small and large insert libraries	2-4 weeks
3	Hand-edited assembly with PCR products incorporated for complete scaffold O&O	3 months

## Next Microbial Targets

### One Genome One Scaffold

- 10x - 3kb shotgun
- Cosmid, Fosmid,  
and Phage End  
Sequencing

DOE RFS

See  
[www.jgi.doe.gov](http://www.jgi.doe.gov)

*Agrobacterium tumefaciens*

*Azotobacter vinlandii*

*Bifidobacterium longum*

*Clostridium thermocellum*

*Dechloromonas aromatica*

*Ehrlichia canis*

*Ehrlichia chafeensis*

*Geobacter metallireducens*

*Haemophilus somnus*

*Lactobacillus brevis*

*Lactobacillus casei*

*Lactobacillus delbrueckii* ssp.  
*bulgaricus*

*Lactobacillus gasseri*

*Lactococcus lactis* ssp. *cremoris*

*Lactococcus lactis* ssp. *lactis*

*Leuconostoc mesenteroides*

*Methanococcoides burtonii*

*Monosiga brevicollis*

*Microbifer degradans*

*Oenococcus oeni*

*Pediococcus acidilactici*

*Pseudomonas syringae* pv.  
*syringae*

*Ralstonia eutropha*

*Streptococcus suis*

*Streptococcus thermophilus*

*Trichodesmium erythraeum*

*Brevibacterium linens*

*Pediococcus pentosaceus*

*Rhodospirillum rubrum*

*Psychrobacter*

*Exiguobacterium*

*Desulfuromonas acetoxidans*

*Desulfovibrio desulfuricans*

Ligated product in pCR4-TOPO

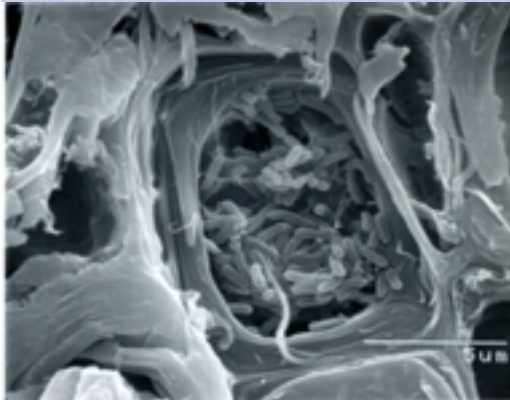
Moss cDNAs

*Ostreococcus*

*Thalassiosira pseudonana* diatom



## *Xylella fastidiosa* an example of sequencing nearest neighbors



Photos by Fundecitrus

- Infects xylem of higher plants, causing a number of economically important diseases
- Different strains are highly related genetically, but have distinct, although often overlapping host specificities.
- Transmitted by the  
glassy-winged  
sharpshooter



## Genome Sequence Status



*Almond Strain*



*Oleander Strain*

Draft sequenced by JGI

Completely sequenced by a Brazilian Consortium



*Citrus Strain*

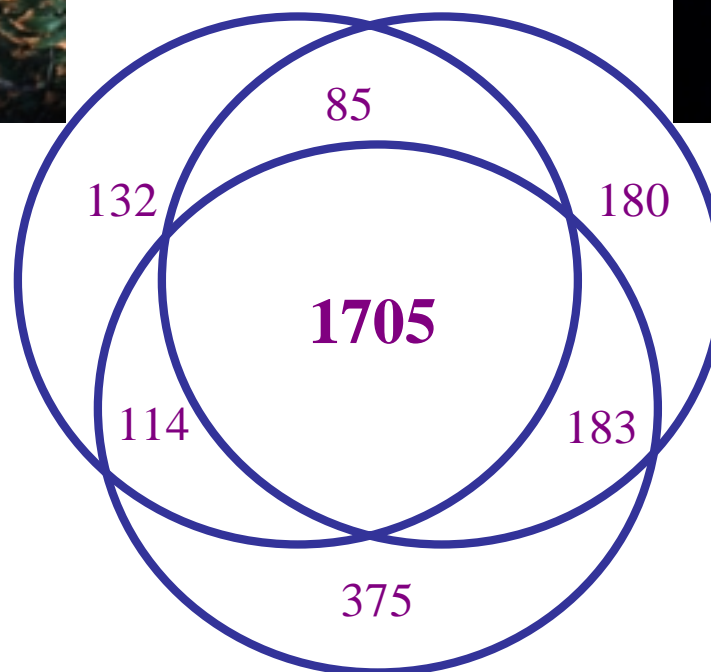
## Gene Repertoire Comparison



*Almond Strain*



*Oleander Strain*



*Citrus Strain*

# Gene Repertoire Comparison

1705 genes in common between all 3 strains, corresponding to ~460 core metabolic and functional pathways.

***This subset of genes contains the main determinants of pathogenicity***

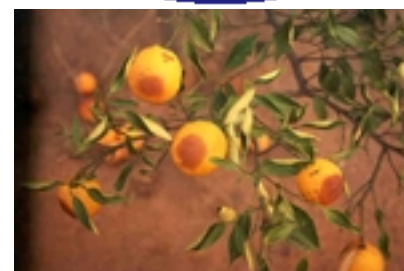
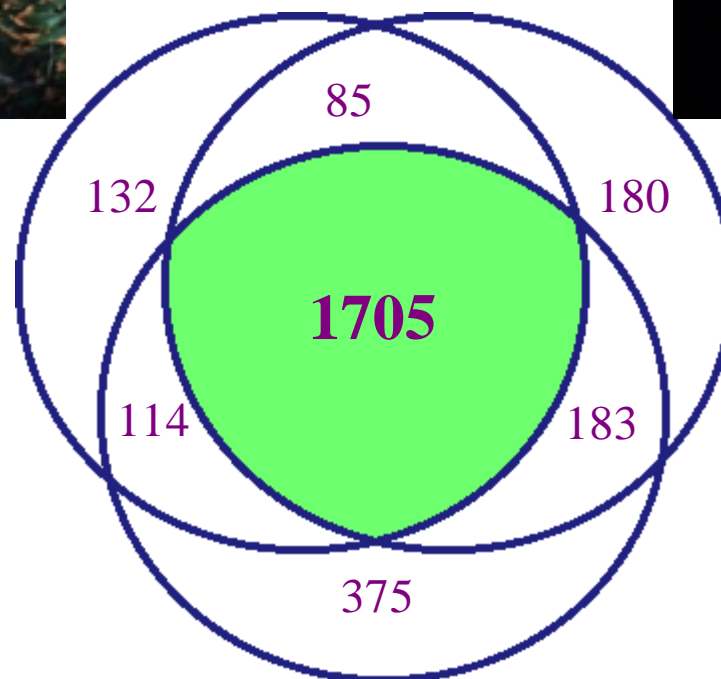
This information can be used to identify functions critical to the life of *Xylella*, and therefore potential drug targets



*Almond Strain*



*Oleander Strain*



*Citrus Strain*

## Unique Genes

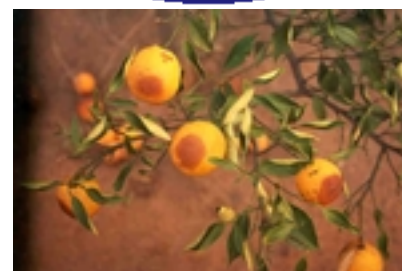
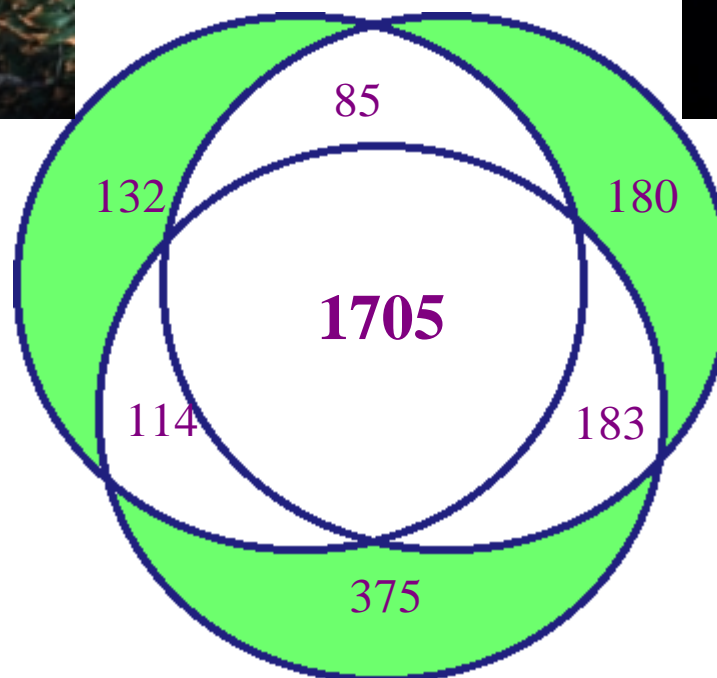
**Genes unique to any specific strain are likely to contain key determinants for host specificity**



*Almond Strain*



*Oleander Strain*



*Citrus Strain*

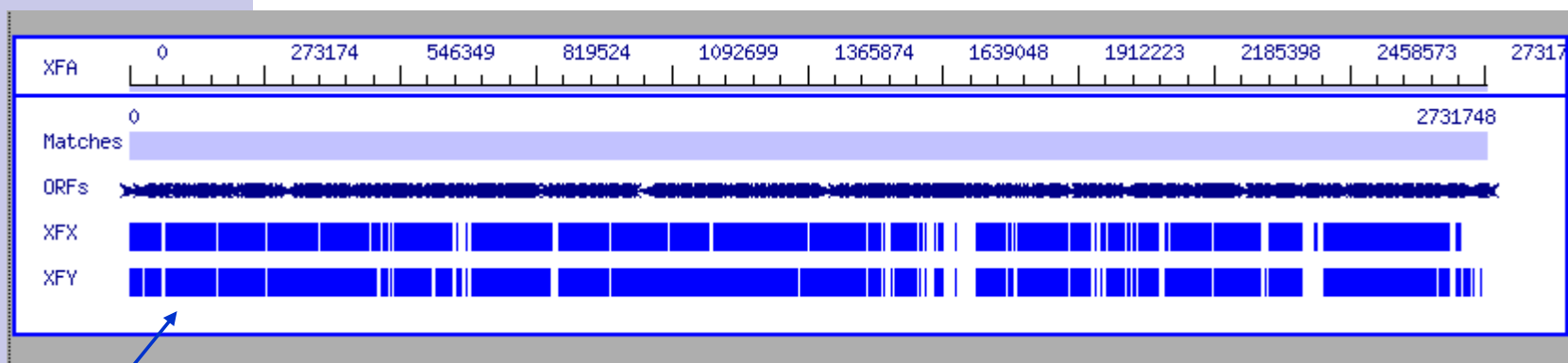


# Genome-Level Gene Comparison

## Strain

Citrus  
(reference)

Almond  
Oleander



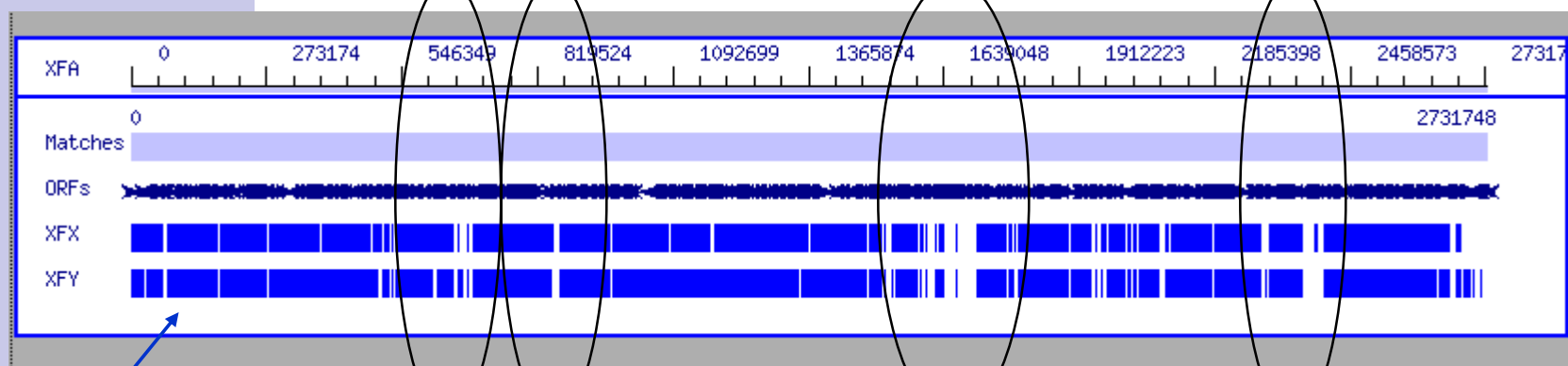
\*blue indicates genes  
shared with reference  
genome

# Genome-Level Gene Comparison

## Strain

Citrus  
(reference)

Almond  
Oleander



\*blue indicates genes  
shared with reference  
genome

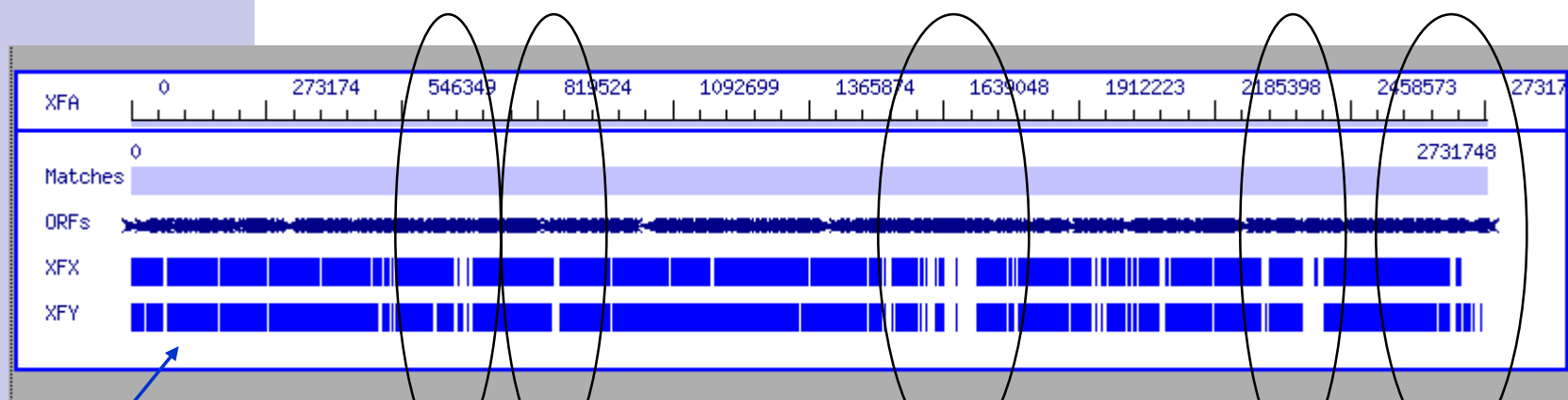
Virus integrated into  
citrus-strain genome

# Genome-Level Gene Comparison

Strain

Citrus  
(reference)

Almond  
Oleander



\*blue indicates genes shared with reference genome

Virus integrated into citrus-strain genome

Plasmid missing in almond strain, partially integrated into genome of oleander strain

## JGI Scientific Programs

- **Illuminating the human genome**
  - Fugu genome sequence
  - Directed mouse sequencing
  - *Human variation*
- **Understanding regulatory Networks**
  - *Ciona intestinalis*
  - *Xenopus tropicalis*
- **Computational algorithms/ tools**
  - Jazz sequence assembler
  - Annotation pipelines (ORNL)
  - Genome Portal / Genome browser

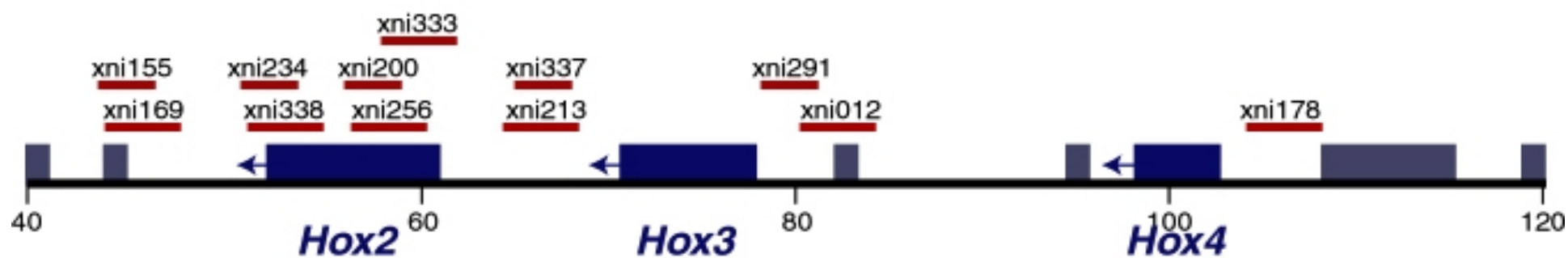
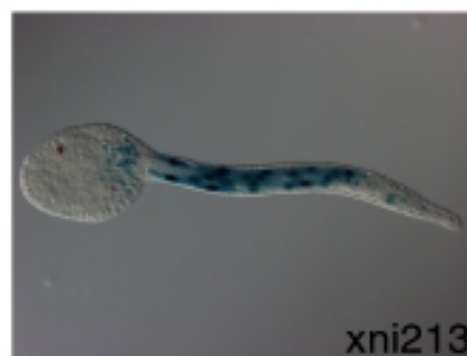
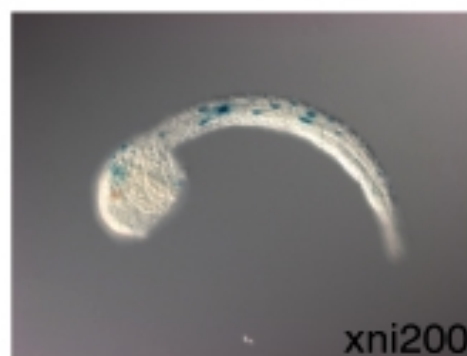
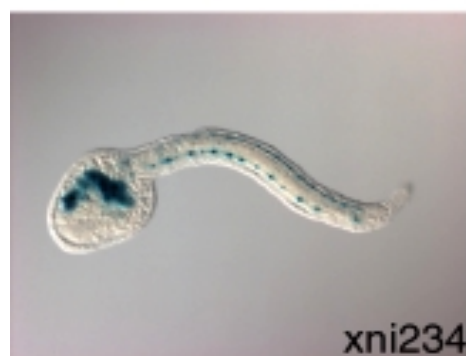
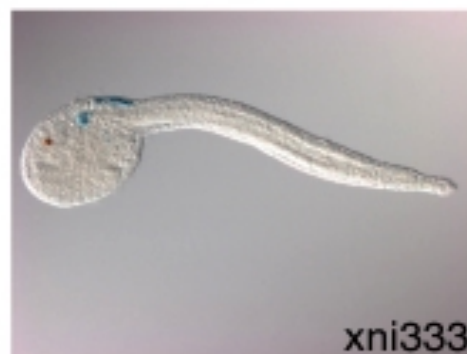
## Fugu Genome Consortium

*Large scale projects*

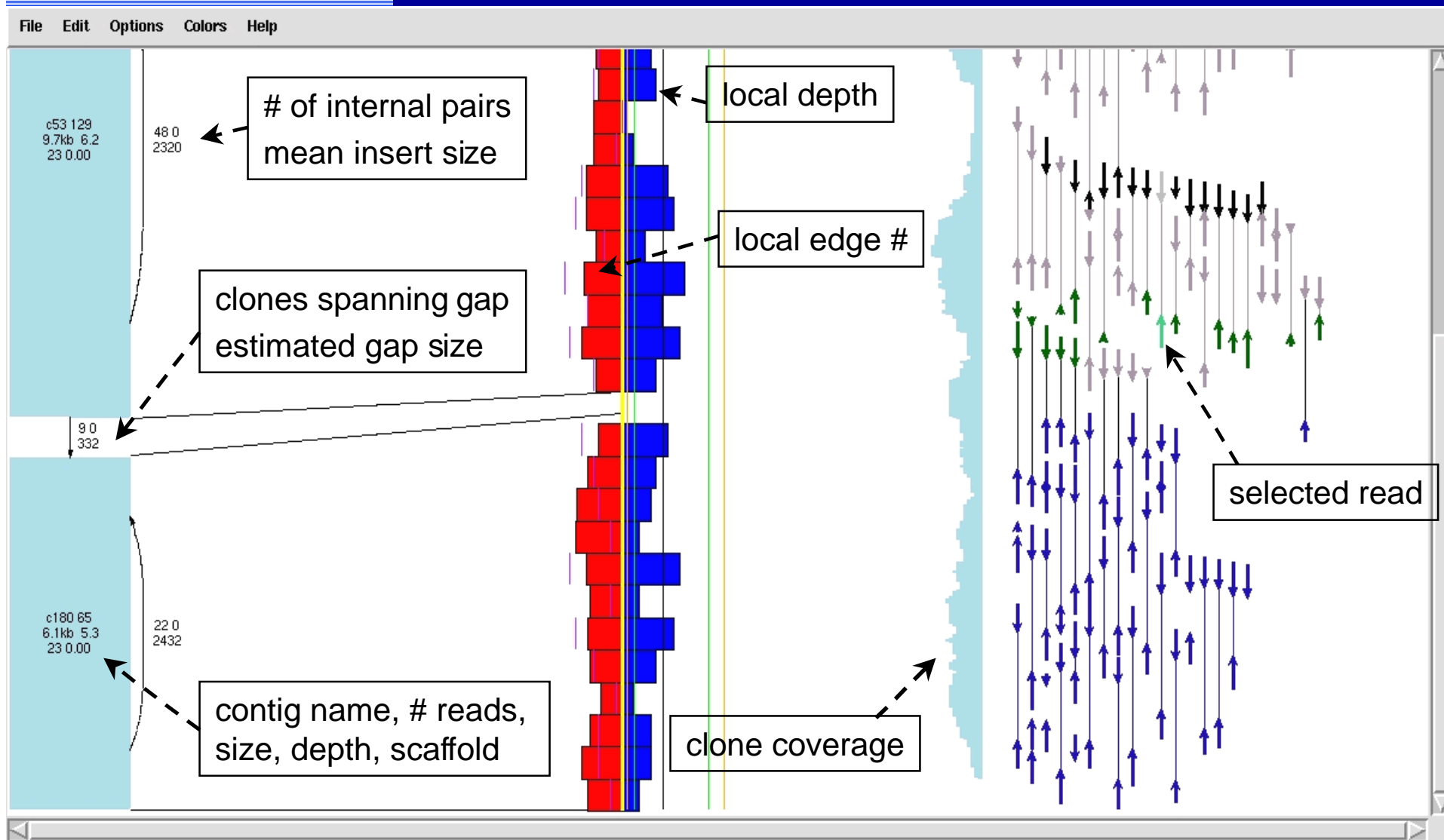
- 365Mb genome
- First vertebrate genome assembled in the public domain after human
- Only example of whole genome assembly in the public domain outside of Celera
- JGI developed JAZZ – whole genome assembler
- Landmark in next phases of the human genome project

- **Formed in December 2000 by JGI**
  - **Members include:**
    - **Joint Genome Institute (lead institute)**
    - **Sydney Brenner**
    - **Lee Hood (Institute for Systems Biology)**
    - **Institute for Molecular and Cell Biology – Singapore**
    - **UK Human Genome Mapping Program – Cambridge**
    - **Myriad & Celera genomics**
- **Current Status: ~5.7X coverage**
- **Recently announced in press release 10-26-01**
- **Data available on [fugu.jgi-psf.org](http://fugu.jgi-psf.org)**

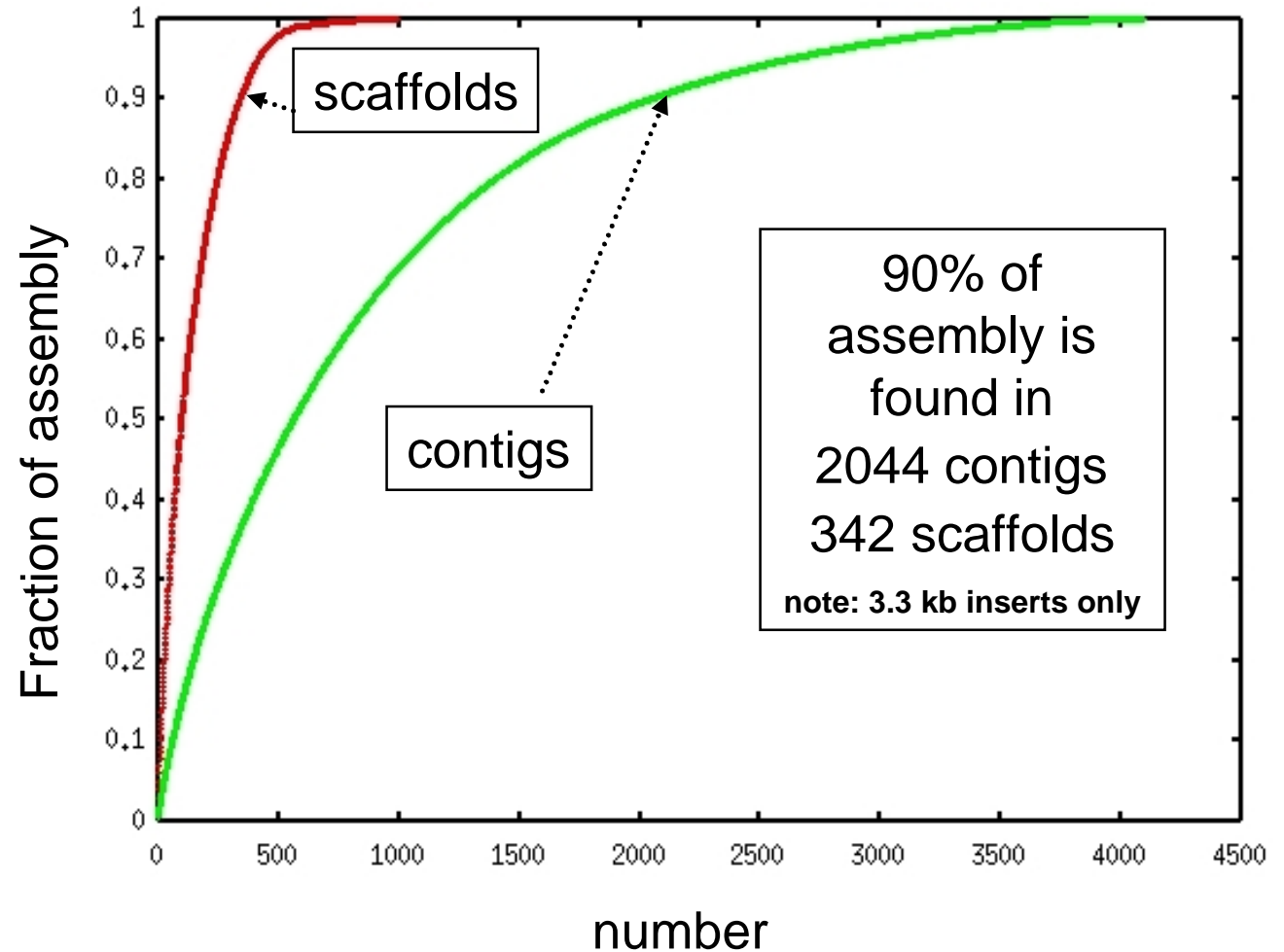
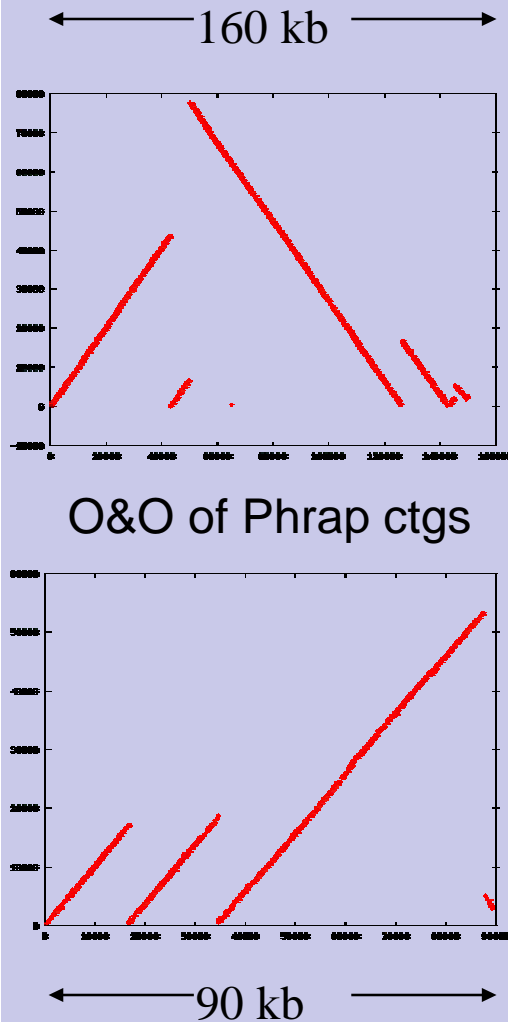




# JAZZ view of assembly



# Phanerochaete chrysosporium (white rot fungus)



## Fugu assembly summary (only 2 and 5.5 kb): 12,403 scaffolds > 2 kb; net length 332.5 MB

Relevant length  
scales:

Typical fugu gene:  
11 kb

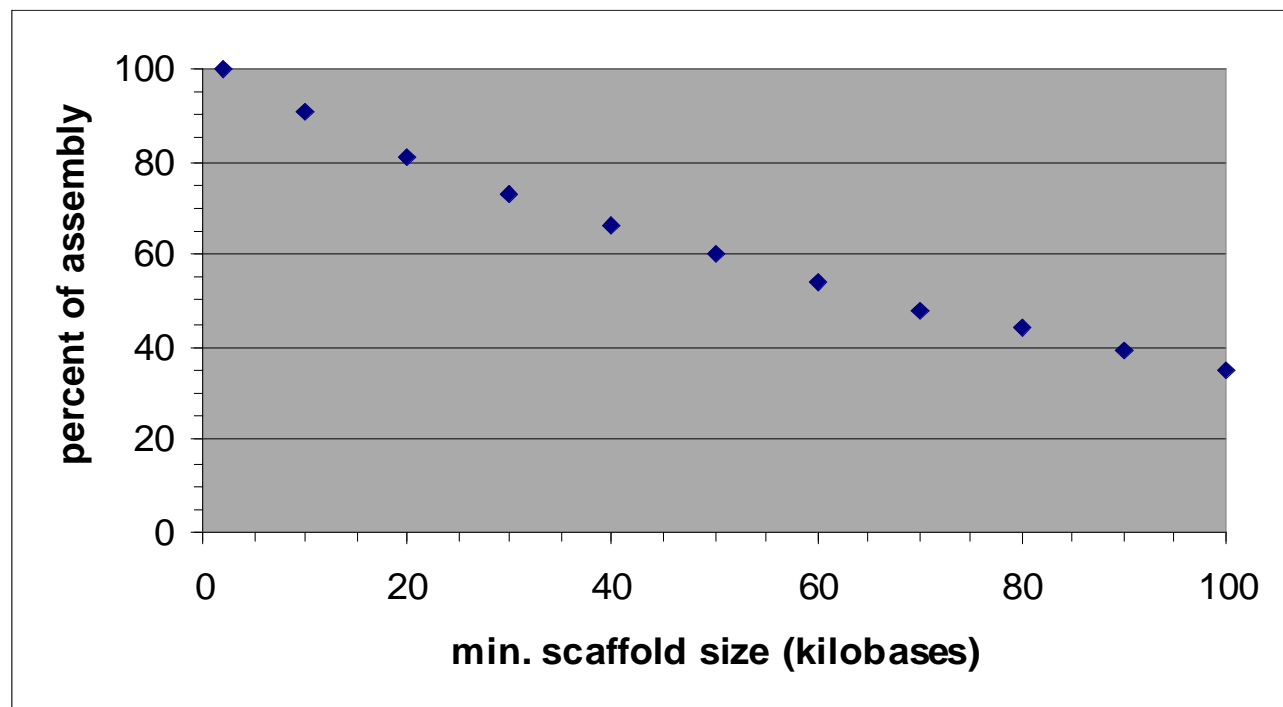
Typical syntenic  
contiguity with  
human genome:  
5 genes = 50 kb  
fugu = 400 kb  
human!

Percent of  
assembled genome  
(euchromatin) in  
scaffolds greater  
than minimum size

**91% of assembly in 6,360 scaffolds > 10 kb (avg 48 kb)**

**81% of assembly in 4,104 scaffolds > 20 kb (avg 66 kb)**

**73% of assembly in 3,067 scaffolds > 30 kb (avg 80 kb)**



**95% of previously known seq. recovered (none misassembled)**

## Fugu Assembly –Map analysis

